American Journal of
EPIDEMIOLOGY

Volume 140
Number 4
August 15, 1994

Sponsored by the Society for Epidemiologic Research

REVIEWS AND COMMENTARIES

Attributable Risk: Advantages of a Broad Definition of Exposure

Sholom Wacholder, Jacques Benichou, Ellen F. Heineman, Patricia Hartge, and Robert N. Hoover

Classification of exposure into two levels—one consisting exclusively of unexposed individuals and the other consisting of exposed and perhaps unexposed ones—yields an unbiased estimate of attributable risk when misclassification is nondifferential. The authors advocate, therefore, the use of a broad definition of exposure when estimating attributable risk. Based on this idea, they justify a simple and robust method for estimating the overall attributable risk from several exposures that is based on a division of subjects into two groups, a baseline consisting of those unexposed to all exposures and everyone else. *Am J Epidemiol* 1994;140:303–9.

bias (epidemiology); biometry; case-control studies; epidemiologic methods; occupational exposure; odds ratio; sensitivity and specificity; statistics

The *attributable risk*, sometimes referred to as the *etiologic fraction* or *attributable risk proportion*, is the fraction of observed cases that would have been avoided if no one in the population were exposed. The attributable risk, just like relative risk, is sensitive to the definition of exposure, but the effects of the definition of exposure are not the same for the measures. In this paper, we review the effects of misclassification on estimates of the attributable risk and propose a strategy that is robust when it is uncertain whether subjects were actually exposed.

The attributable risk decreases with the lower relative risk resulting from misclassification of unexposed subjects as exposed, but increases with the proportion considered exposed. Thus, the effects of alternative classification strategies are not immediately clear. In this paper, we examine the effects of misclassification and of categorization of exposed and unexposed subjects on the attributable risk and discuss the problem that stimulated our thinking about this issue. We demonstrate that the attributable risk is robust to the inclusion as exposed of subjects who are unexposed and discuss the impact of adjusting for additional risk factors. We justify a

Received for publication August 16, 1993, and in final form May 6, 1994.

Abbreviations: NOHS, National Occupational Hazard Survey.

From the Epidemiology and Biostatistics Program, National Cancer Institute, Rockville, MD.

simple and assumption-free method for calculating the overall attributable risk due to more than one variable. We conclude that when calculating attributable risk, investigators should strive for a definition of exposure such that everyone in the unexposed group has no more risk than the truly unexposed, while the exposed group may consist of exposed and unexposed individuals.

EXAMPLE

Questions regarding attributable risk arose from a study of mesothelioma and asbestos. Our example is based on data from a case-control study by R. Spirtas, E. F. Heineman, L. Bernstein, et al., "Malignant Mesothelioma: Attributable Risk of Asbestos Exposure" (unpublished manuscript, submitted for publication). Analyses based on a subset of these data are presented here to make our methodological point; readers interested in substantive questions will need to await publication of the paper. All participants in the study were assigned an estimated probability of occupational exposure to asbestos by evaluation of their complete work history and by comparison of industry/occupation combinations in the work history with the exposures identified in the National Occupational Hazard Survey (NOHS) (1). Individuals were classified into one of four categories of likelihood of asbestos exposure: 0, extremely low (no known exposure); 1, moderately low (NOHS probability of exposure: >0 percent to 19 percent); 2, medium (NOHS probability of exposure: >20 percent); and 3, high (insulators, shipyard workers, furnace or boiler installers and repairers). The NOHS classification was overridden to create the "high" category because it was believed that the NOHS classification was not completely germane for the period of the 1940s through the 1960s, when most of the participants were employed. Individuals who cohabited with asbestos-exposed workers, lived near an asbestos mine or mill, or who claimed exposure to asbestos were placed in the "moderately low" category, even if their job history gave no evidence of asbestos exposure.

Table 1 shows the classification of cases and controls by assigned likelihood of asbestos exposure and family history, a dichotomous variable indicating whether one or more first degree relatives had had a diagnosis of cancer. Table 2 shows the crude level-specific odds ratios and attributable risks according to the estimated likelihood of asbestos exposure. Note that, as expected, the odds ratio increases with the estimated probability of exposure. This gradient may reflect not only increased probability of exposure, but also, to some extent, differences in the average levels or durations of exposure in persons who actually were exposed in each category.

Table 3 shows how the crude estimate of the odds ratio and the attributable risk change when the cutoff for classification as exposed changes. For example, using the broadest criterion of exposure results in comparisons of categories 1–3 with category 0 and an attributable risk of 0.82, exactly equal to the sum of the three level-specific attributable risks in table 2. In the

TABLE 1. Mesothelioma cases and controls according to likelihood of exposure and family history*

Exposure class	Likelihood of exposure	Family history		No family history	
		No. of cases	No. of controls	No. of cases	No. of controls
0	Extremely low	3	37	8	108
1	Moderately low	16	56	17	76
2	Medium	7	9	5	16
3	High	53	41	74	80
Total		79	143	104	280

* Based on data from a case-control study by R. Spirtas, E. F. Heineman, L. Bernstein, et al.: "Malignant mesothelioma: attributable risk of asbestos exposure" (unpublished manuscript).

TABLE 2. Level-specific attributable risks according to likelihood of exposure*

Exposure class	Likelihood of exposure	Disease status: No. of cases	Disease status: No. of controls	Proportion among controls	Odds ratio	Level-specific attributable risk†
0	Extremely low	11	145	0.34	1.0	0.0
1	Moderately low	33	132	0.31	3.3	0.13
2	Medium	12	25	0.06	6.3	0.06
3	High	127	121	0.29	13.8	0.64
1, 2, 3	Non-trivial	172	278	0.66	8.2	0.82

* Based on data from a case-control study by R. Spirtas, E. F. Heineman, L. Bernstein, et al.: "Malignant mesothelioma: attributable risk of asbestos exposure" (unpublished manuscript).

† Fraction of the total cases avoided if a subject at a given likelihood of exposure had the same risk as those at extremely low likelihood of exposure.

TABLE 3. Effect of changing the criterion for exposure*

Exposed categories	Unexposed categories	Proportion of exposure among controls	Odds ratio	Attributable risk
1, 2, 3	0	0.66	8.2	0.82
2, 3	0, 1	0.35	6.0	0.63
3	0, 1, 2	0.29	5.7	0.57

* Based on data from a case-control study by R. Spirtas, E. F. Heineman, L. Bernstein, et al.: "Malignant mesothelioma: attributable risk of asbestos exposure" (unpublished manuscript).

example, the attributable risk decreases as the criterion for classification as exposed is made more stringent, decreasing the proportion exposed and the odds ratio estimate.

We also examined the effect of misclassifying subjects called "unexposed" as exposed, mimicking what might occur under broad definitions of exposure. In table 4, we calculate the crude attributable risks when everyone with moderately low or higher risk is called "exposed," while moving increasingly larger subsets of subjects from the baseline exposure group to the exposed group nondifferentially by disease status. Note that, while the odds ratio decreases with the amount of misclassification of "unexposed" into "exposed," the calculated attributable risk under each scenario always equals the actual attributable risk, though the standard error of the estimate (2) does increase. Thus, classifying exposed subjects as unexposed (table 3) reduces the estimate of attributable risk, while classifying unexposed subjects as exposed (table 4) does not change it. Either type of error, of course, reduces the calculated odds ratio.

PROPERTIES OF THE ATTRIBUTABLE RISK

Attributable risk has a distributive property: the sum of the attributable risks from two or more categories of exposure equals the attributable risk calculated from combining those categories into a single exposed category, regardless of the number and the divisions of the categories that are formed (2, 3). Thus, the attributable risk from com-

TABLE 4. Effect of misclassifying unexposed subjects as exposed*

Proportion at level 0 misclassified as exposed	Proportion of exposure among controls	Odds ratio	Attributable risk	Standard error†
0.0	0.66	8.2	0.82	0.053
0.20	0.73	7.5	0.82	0.059
0.40	0.79	6.9	0.82	0.069
0.60	0.86	6.5	0.82	0.085

* Based on data from a case-control study by R. Spirtas, E. F. Heineman, L. Bernstein, et al.: "Malignant mesothelioma: attributable risk of asbestos exposure" (unpublished manuscript).

† Standard error of the attributable risk.

bining exposure classes 1, 2, and 3 is 0.82, exactly equal to the sum of the attributable risks from the three separate categories (0.13 + 0.06 + 0.64) (table 2). Further, wrongly classifying unexposed subjects as exposed does not affect the value of the attributable risk when there is a pure baseline group of unexposed subjects and nondifferential misclassification, that is, the probability of classifying an unexposed subject as exposed is the same for cases and controls (4). By the distributive property, the attributable risk for the combined group of both truly exposed and of presumed but not actually exposed subjects is the sum of the attributable risks from the two subgroups. Since the attributable risk from the unexposed subgroup is zero by definition, the attributable risk for the combined group must equal the attributable risk from the exposed only. Thus, as in table 4, the ratio of the numbers of exposed to unexposed individuals in the combined category does not affect the attributable risk. The estimate of attributable risk is unbiased as long as all exposed individuals are classified as exposed, regardless of the proportion of unexposed subjects who are misclassified nondifferentially as exposed. Therefore, a sensitive classification scheme is an appropriate strategy, even when specificity is exceedingly low.

The increase in the estimate of attributable risk with an increasingly broad definition of exposure will hold as long as each new set of subjects added to the definition of exposed has an estimated relative risk of greater than 1, even when the result is a decreasing odds ratio of exposed relative to unexposed subjects. A relative risk below 1 for the new set included as exposed would increase the proportion exposed, but reduce the odds ratio and the attributable risk. A relative risk of exactly 1 for the new set included as exposed will leave the attributable risk unchanged. Thus, truly unexposed subjects, nominally exposed subjects who are not actually exposed (as a result of exposure misclassification based on job title or of protective equipment), and truly exposed subjects who are at no greater risk than the unexposed (perhaps as the result of a mitigating genetic factor) can all be included with the exposed.

IMPLICATIONS

In our example, the attributable risk is unaffected when unexposed cases and controls are misclassified as exposed nondifferentially and exposed ones are classified correctly. We also made a heuristic argument for this fact. A general proof of this fact is found in reference 4. We conclude that use of the broadest definition of exposure provides the best estimate of attributable risk.

These properties have implications for defining exposure for purposes of calculation of the attributable risk when misclassification is nondifferential. First, exposed individuals must be classified as exposed. Using a more restrictive criterion for exposure will reduce the estimate of attributable risk, as shown in table 3. Second, misclassifying unexposed individuals as exposed does not affect the point estimate of the attributable risk, but does reduce precision. In table 4, the attributable risk does not change when increasing numbers of unexposed subjects are classified as exposed. Therefore, in the study of Spirtas et al., subjects who claimed exposure to asbestos were included as exposed, even if it appeared unlikely from their work history alone. Third, careful delineation of the levels of exposure in the exposed is also unimportant. As in table 2, the distributive property ensures that any grouping of exposed subjects does not affect the estimate of attributable risk. In summary, any reasonable rule that correctly classifies nearly all exposed subjects can give a good estimate of attributable risk, regardless of the proportion of unexposed subjects who are wrongly classified as exposed. However, failure to include the "moderately low likelihood" of exposure category, or any other category that has increased risk that should be attributed to exposure, would underestimate the impact of exposure on disease in the population. These considerations do not apply to estimates of relative risk, which generally will be distorted unless sensitivity and specificity are 100 percent.

There is a cost in precision associated with misclassifying unexposed subjects as exposed, even for the attributable risk, as seen in table 4. Equation 7 of reference 2 implies that the standard error of the attributable risk estimate increases as the proportions of exposed cases and controls increase.

EFFECT OF ADJUSTMENT FOR ADDITIONAL RISK FACTORS

Does our argument apply when estimating attributable risk while adjusting for another risk factor? We have described the advantages of a broad definition in the univariate situation. Moreover, when the attributable risk is calculated on the basis of a saturated model (one that includes all possible interactions among the variables in the model), the distributive property also holds (2), and use of a broad definition of exposure that includes unexposed subjects will have no effect on the estimate, although the estimate of its variance will increase, as in our example. However, in an unsaturated model that includes effects of other risk factors, the distributive property may not apply exactly, and thus, the specificity of exposure can have a small effect on attributable risk, even when the sensitivity is 1.0.

For the mesothelioma data set, we calculated the level-specific attributable risks for each of three levels of likelihood of occupational exposure and attributable risk based on a dichotomization of exposure into trivial or nontrivial likelihood. In the univariate analysis (table 2), the sum of the three level-specific attributable risks was 82.46 percent, which, as expected, was exactly equal to the overall attributable risk. We then used the method of Bruzzi et al. (5) to adjust for dichotomous family history of cancer, one of the variables most strongly related to mesothelioma among possible risk factors in the data set, in an unsaturated (no interaction) model. The sum of the level-specific attributable risks was 81.91 percent; the attributable risk based on dichotomous exposure was 82.09 percent. When interactions between occupational and family history were included (saturated model), both were 82.33 percent.

This example is consistent with our experience that the distributive property is a reasonable approximation in the more usual situation of a model with several risk factors in addition to the exposure, unless one of these risk factors also acts as a strong effect modifier of the exposure. We therefore believe that, as a practical matter, our points also apply in the presence of confounding.

OVERALL ATTRIBUTABLE RISK FROM SEVERAL RISK FACTORS

Coughlin et al. (6) have suggested that assuming an additive relation between covariates included in a logistic model can yield an overall estimate of the attributable risk due to the joint effects of the risk factors. One application of our work is the justification of a simple procedure used by Hartge et al. (7) to estimate the total attributable risk due to two or more risk factors without requiring any modeling assumptions. In this method, subjects are divided into two categories—1) not exposed to any of the risk factors or 2) exposed to at least one of the risk factors—and the attributable risk for this dichotomous variable is calculated in the usual way. This procedure will lead to an unbiased estimate of the risk attributable to either one or both of the risk factors, as long as no one called unexposed is truly exposed to any risk factor. However, the estimate of the joint attributable risk from several common exposures may be imprecise if the size of the baseline group is small.

Example

For the mesothelioma data set, we were able to estimate the joint attributable risk from asbestos exposure (defined as the nontrivial likelihood of exposure) and family history of cancer. The proportion of controls with either asbestos exposure or family history of cancer is 315/423 = 0.74. The odds ratio for those who reported at least one of these as compared with the proportion that had neither asbestos exposure nor family cancer was 7.5, yielding an attributable risk of 82.9 percent. The adjusted attributable risk for occupational history alone was 82.1 percent, indicating that the impact of family history would be very small after eliminating asbestos exposure, even though the adjusted attributable risk for family history is 12.3 percent.

DISCUSSION

The attributable risk will not automatically be higher when a broad rather than a restrictive definition of exposure is used. For example, Vineis et al. (8) found an attributable risk estimate of 10.9 percent for lung cancer in subjects born before 1930 working in an a priori list of "industries/occupations with well-established carcinogenic exposures," but a lower attributable risk of 8.8 percent when a broader definition that also includes workers in "industries/occupations with suspect carcinogenic exposures" was used. The most likely explanation of their findings is that workers in some of the occupations believed to be hazardous had, in fact, reduced risk of disease in their study. One might argue, therefore, that one should exclude from the exposed group all individuals in suspect occupations whose observed relative risk is less than 1.0. This strategy would tend to bias the estimate of the attributable risk upward, since occupations with randomly low observed relative risks might be excluded, while those with randomly high observed relative risks would be included at the observed level. Random error leading to exposure categories falsely being associated with disease risk or falsely found to be unrelated to risk can create misleading estimates of attributable risk when a data-based criterion of "exposed" (such as workers in any occupations with odds ratios greater than 1.1) is used.

Of course, there are circumstances in which a broad definition confers no advantage. Sometimes, the separated effects of various levels of exposure can reveal the likely impact of reduction rather than elimination of the exposure. Also, when rates of misclassification of exposure differ greatly between cases and controls, neither a broad nor a restrictive definition of exposure can ameliorate the problem. While we have no direct evidence from the asbestos study of Spirtas et al. that supports or refutes the nondifferentiality assumption, a violation seems possible, since asbestos is a well-known risk factor for mesothelioma and next-of-kin respondents for some of the cases may have exaggerated the likelihood of exposure. If cases falsely claim exposure more often than controls, the specificity of exposure for cases will be lower than that for controls, and the estimate of attributable risk, as well as of relative risk, will be biased upward. Finally, there is loss of precision when the definition of exposure encompasses levels of exposure that have the same risk of disease as the unexposed (table 4).

The attributable risk is one measure of the burden of disease that is caused by an exposure. We have pointed out that estimates of the attributable risk, unlike ratios and differences of risk between exposed and unexposed persons, are robust to lack of specificity in detecting exposure. Nonetheless, there are well-known problems of interpretation with attributable risk. It depends on the exposure distribution in the studied population, which makes it likely to vary from community to community (3). Also, it does not distinguish between two situations with very different implications regarding intervention strategies: one with many individuals at small excess risk and one with few individuals at high excess risk. The relative risk estimate can help to distinguish between these situations.

REFERENCES

1. National Occupational Hazard Survey. Vol III. Survey analysis and supplemental tables. Cincinnati, OH: National Institute of Occupational Safety and Health, 1977. (DHEW publication no. (NIOSH)78–114).
2. Benichou J. Methods of adjustment for estimating the attributable risk in case-control studies: a review. Stat Med 1991;10:1753–73.
3. Walter SD. The estimation and interpretation of

attributable risk in health research. Biometrics 1976;32:829–49.
4. Hsieh CC, Walter SD. The effect of non-differential exposure misclassification on estimates of the attributable and prevented fraction. Stat Med 1988;7:1073–85.
5. Bruzzi P, Green SB, Byar DB, et al. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol 1985;122:904–14.
6. Coughlin SS, Nass CC, Pickle LW, et al. Regression methods for estimating attributable risk in population-based case-control studies: a comparison of additive and multiplicative models. Am J Epidemiol 1991;133:305–13.
7. Hartge P, Harvey EB, Linehan WM, et al. Unexplained excess risk of bladder cancer in men. J Natl Cancer Inst 1990;82:1636–40.
8. Vineis P, Thomas T, Hayes RB, et al. Proportion of lung cancers in males, due to occupation, in different areas of the USA. Int J Cancer 1988; 42:851–6.